AREDS dbGaP Data Tables: A User's Guide

Version 1.0

TABLE OF CONTENTS

1. Description of the AREDS dbGaP Data Tables 1

1.1. Project Data Tables 2

1.1.1. Tables that contain a single record for each participant 3

1.1.1.1. Eye Disease Phenotypes [table name: amdlensphenotype]. 3
1.1.1.2. Enrollment/Randomization [table name: enrollment_randomization]. 3
1.1.1.3. Mortality [table name: mortality] 4

1.1.2. Tables that contain multiple records for some or all participants 5

1.1.2.1. Follow-up (all participants) [table name: followup] 5
1.1.2.2. NSAIDs Use (all participants) [table name: nsaids] 6
1.1.2.3. Fundus Photographs (all participants) [table name: fundus] 7
1.1.2.4. Lens Photographs (all participants) [table name: lens] 9
1.1.2.5. Visual Functioning Questionnaire [VFQ-25] [table name: vfq] 10
1.1.2.6. Dietary Information [table name: dietary] 11
1.1.2.7. Adverse Experiences (all participants) [table name: adverse] 11
1.1.2.8. Hospitalization [table name: hospitalization] 12

1.2. Phenotype Data Tables for Participants with a Genetic Specimen 12

1.2.1. Eye Disease Phenotypes [table name: genspecphenotype] 13
1.2.2. Fundus & Lens Photographs [table name: genspecfunduslens] 15

2. Analysis of AREDS Data 15

2.1. Characterizing the Cohort 15

2.1.1. Randomized Treatment (from Enrollment/Randomization table) 15
2.1.2. Clinical Visual/Ocular Characteristics (from Enrollment/Randomization table) 16
2.1.3. Eligible Study Eyes for AMD Clinical Trial (from Fundus Photographs table) 16
2.1.4. Photographic Ocular Characteristics (from Fundus or Lens Photographs table) 16
2.1.5. Dietary Characteristics 17

2.2. Determining Outcomes 17

2.2.1. Ocular Outcomes 17
2.2.2. Other Non-ocular Outcomes 18
2.2.3. Clinical Trial vs. Natural History Phases of AREDS 18

AREDS dbGaP Data Tables: A User's Guide

Analysis of the AREDS data being provided on the NCBI dbGaP is not trivial and prone to user error if the user is not familiar with both the study design and how the variables are defined. This User's Guide is designed to help researchers perform meaningful analysis on the AREDS data. It is divided into two parts:

- Part I is a description of the AREDS dbGaP data tables.

- Part II describes key information needed to do the data analysis, such as how participants were randomized, how ocular outcomes were defined, and what eyes were considered eligible study eyes.

1. Description of the AREDS dbGaP Data Tables

There are two groups of data tables:

- Project Data Tables:
  - 11 data tables are included.
  - Data from all participants enrolled in AREDS are included.
  - ID is used as the subject identifier.

- Phenotype Data Tables:
  - 2 data tables are included.
  - Only data from participants with a genetic specimen are included.
  - ID2 is used as the subject identifier.

If a participant consented to their genetic specimen being for General Research Use, ID and ID2 identify the same person. However, a participant who consented to their specimen being used for Eye Disease Research Only has a different ID2 number which is a 4-digit code that begins with the letter 'G'.

Other general comments about the data tables are as follows:

- "Phase II" implies the AMD and lens opacity Clinical Trial phase, which ran from November 1992 to October 2001. "Phase III" implies the Natural History phase, which continued after the clinical trial concluded. Phase III ended in December 2005.

- "Time" fields on each table were measured as the number of years between date of randomization and date of form completion, or number of days between date of event and date of resolution.

- There were two different baseline visits: the qualifying visit and the subsequent randomization visit which were typically 1-3 months apart. Fundus photographs were taken at the qualifying visit. Once these photographs were graded, if the participant was eligible for the study, then they returned to the clinic for the randomization visit.

- Participants determined to be in AMD Category 1 at enrollment were randomly assigned to placebo or antioxidants with equal likelihood. Otherwise, participants were randomly assigned to one of the four treatment groups with equal likelihood.

- The study visit number (VISNO) is included in multiple tables (FOLLOWUP, NSAIDS, FUNDUS, LENS, and VFQ) as a way to link information between these tables at corresponding time points. Each visit number represents an approximate 6-month interval after randomization (i.e., 01 = 6-month visit after randomization, 02 = 12-month visit after randomization, etc.).

- In Phase II, visits 01, 03, 05, etc. were considered "non-annual" (or odd-numbered) visits, whereas visits 02, 04, 06, etc. were considered "annual" (or even-numbered) visits. Additional information was gathered at annual visits and is thus missing by design at non-annual visits. At the end of Phase II, one odd-numbered visit was allowed to be an annual visit, as additional useful information could be gathered just prior to the end of the clinical trial.

- In Phase III, participants were typically seen once per year, and the data collected from this visit (similar to the annual visit in Phase II) were entered on an In-clinic Visit form. The clinics had the option of collecting this data at an even-numbered visit (which was preferred) or collecting this once-per-year data at an odd-numbered numbered visit in cases where the participant was not available to schedule an even-numbered visit in person. Data from the interim 6-month visit were collected via a telephone call and entered onto the Contact form. The Contact visit could also be even-numbered or odd-numbered.

Note:

Further detail on specific variables can be found in the data dictionary for each table, which is available in the Publicly Available Data download area:

ftp://ftp.ncbi.nlm.nih.gov/dbgap/NEI/AREDS/phs000001.v2.p1

The data dictionaries for each table are considered essential reading in conjunction with this document. Variables listed in the data dictionaries that are not directly addressed in this document correspond with a question on a case report form and therefore do not require further explanation. Case report forms can be seen by clicking on the 'Documents' tab and looking in the 'Case Report Forms' folder on the right-hand side of the page:

http://www.ncbi.nlm.nih.gov/projects/gap/cgi-in/document.cgi?study_id=phs000001.v2.p1&phv=173&phd=1&pha=2&pht=371&phsf=0&phvf=&phdf=&phaf=&phtf=&dssp=1

1.1. Project Data Tables

These 11 tables include the essential data collected during AREDS as well as final phenotype data for **all participants**. The tables can either contain a single record for each participant or multiple records for each participant. All Project Data Tables use the ID field as the unique participant identifier.

1.1.1 Tables that contain a single record for each participant

1.1.1.1. Eye Disease Phenotypes [table name: amdlensphenotype].

This table primarily shows the Final AMD and Final Cataract phenotype status, but also indicates whether the participant provided a genetic specimen. There is a single Final AMD phenotype variable, whereas there are multiple Final Cataract phenotype variables.

Table details are as follows:

- AMDSTAT: A detailed description of how the final AMD phenotype categories shown in this field were defined is found at the following link:

  http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd001138.1

- Most of the remaining fields in this table are related to the Final Cataract phenotype. A detailed description of how the cataract phenotype was defined is found in the following article:

  AREDS Report No. 24 – Cataract Classification using Serial Examinations in the Age-Related Eye Disease Study. *American Journal of Ophthalmology*. Vol 145 (No. 3), March 2008. pp 504-508.

- LCORSCORE, RCORSCORE, LPSCSCORE, RPSCSCORE: These fields are opacity percentages and as such you would expect that the maximum value is 100%. However, since these fields are predicted based on regression models, it is possible that the prediction will produce a value that is over 100%.

- HASGENSP: This indicates whether the participant provided a genetic specimen. However, not all specimens currently have DNA available. In order to see if a specimen has DNA available for ordering, please visit the NEI-AREDS Genetic Repository web catalog:

  http://ccr.coriell.org/Sections/Collections/AREDS/DNAs.aspx?PgId=579&coll=ED

1.1.1.2. Enrollment/Randomization [table name: enrollment_randomization].

This table contains a broad spectrum of characteristics captured at the time the participant enrolled in AREDS including, but not limited to:

Demographics: gender, age, race, education
Health history: blood pressure, cancer, smoking, angina, diabetes
Current supplementation and medication use
Prior eye treatments
Vision status
AMD Category (see AREDS Manual of Procedures section 3.1.2.8)
Study treatment assignment (Placebo, Antioxidants, Zinc, or Antioxidants+Zinc)

Table details are as follows:

- There were two different enrollment visits: the qualifying visit and the subsequent randomization visit that were typically 1-3 months apart. Fundus photographs were

taken at the qualifying visit. Once these photographs were graded, study participants found to be eligible for the study returned to the clinic for the randomization visit.

- Participants determined to be in AMD Category 1 at enrollment were randomized to placebo or antioxidants with equal likelihood. Otherwise, participants were randomized to one of the four treatments with equal likelihood.

- Vitamin usage (MULTIVCU, VAIUCURR, VARECURR, BCIUCURR, BCMGCURR, VITCCURR, VEIUCURR, VEMGCURR, ZINCCURR, CALCCURR, IRONCURR, SELECURR fields): These fields indicate participant's use current at the time of enrollment, rather than usage beyond enrollment. If use is current, length of use is specified. A response of 'N' indicates that use of that vitamin occurred in the past but not currently, whereas a response of 'X' indicates no prior use of that vitamin.

- Centrum usage (CNTRMTAK): This field indicates the participant's intent to take the multivitamin Centrum® during AREDS.

- Two blood pressure measurements (diastolic/systolic) were taken during the enrollment interview: one at the beginning of the interview and a second later in the interview.

- Visual acuity at the initial (qualifying) visit (QUALVARE, QUALVALE) was used to determine AMD Category, but otherwise disregarded. At the qualifying visit, vision from Chart R was used if the acuity in both eyes was 74 letters or higher; otherwise best-refracted visual acuity was obtained from Charts 1 & 2. When an eligible participant was randomized into the study, the best-refracted visual acuity at randomization from Charts 1 & 2 (ACUIT1RR, ACUIT2RL) was used as the baseline visual acuity.

- Refractive error at the randomization visit (RE_ERR_R, LE_ERR_R) was captured for all participants. Refractive error at the qualifying visit (RE_ERR_Q, LE_ERR_Q) was captured only if visual acuity in either eye was <74 letters when utilizing Chart R.

- HYDROCHL, DIPYRIDA, TRIAMTER, SPIRONOL, PROPRANO, BETABLOC, CHLORTHA, FUROSEMI, ACEINHIB, CALCBLOC, POTASSIU, DIGOXIN, NITROGLY, ISOBDINI, ALLOPURI, BENEMID, BLOODTHN, LOWRCHOL, ANTIINFL, THYRHORM: These medication use measures indicate number of years taken for any medication that the participant is currently taking at the time of enrollment.

- BMI_R: body-mass index calculated from height and weight measures from baseline data.

1.1.1.3. Mortality [table name: mortality]

This table contains select information regarding participants who died during AREDS (i.e., between November 1992 and December 2005), including: length of follow-up until death, cause of death (using ICD coding), and degree of reliability that the death occurred.

Deaths were documented during AREDS on a Death Report (through the end of 2005). After the conclusion of AREDS, a death search was conducted for all AREDS participants using the National Death Index (NDI). The results provided by NDI include a CLASS variable (which ranges from 1 to 5 with a 1 being the best) that is the level of assurance that the person whose death information was requested is the same as the person whose information is being provided by the NDI. In some cases, deaths recorded on an AREDS Death Report were assigned a CLASS value by NDI that was not equal to 1. However, deaths reported during AREDS should be assumed to have occurred, regardless of the value of the CLASS variable reported by NDI. For deaths only reported by the NDI search, it is at the discretion of the researcher whether to include them in any mortality analyses based on the value of the CLASS variable.

Table details:

- Cause of death was determined either broadly by the study coordinator (CAUSE), more specifically by a certified ICD-9 coder (ICD9COD1), or via the NDI search (ICD10). In some cases, one, two or all of these fields may be missing.

- AREDSDTH: Indicates whether the death was discovered during the AREDS study (through 2005), or subsequently by the NDI search.

1.1.2 Tables that contain multiple records for some or all participants

1.1.2.1 Follow-up (all participants) [table name: followup]

This table contains information collected at routine subjects visits, including, but not limited to:

Health status: Cancer, diabetes, stroke, vascular conditions, arthritis, gout
Clinical Ocular status: Glaucoma, retina & lens treatments
Refractive error
Current supplementation and drug use
Smoking status
Blood pressure
Body-mass index
Study treatment

Table details are as follows:

- Routine visits occurred every 6 months during AREDS. Visit numbers represent the 6-month visit (i.e., 01 is the 6-month visit, 02 is the 12-month (1-year) visit, etc.).

- Some follow-up information was only collected annually (e.g., visits 02, 04, 06, etc.) and thus is missing by design at certain visits (e.g., visits 01, 03, etc.).

- Some follow-up information was collected less frequently than annually - only at visits where the Follow-up Interview occurred (i.e., visits where FOLUPINT is not missing). This interview typically occurred at year 5, at the end of the clinical trial (typically year 6, 7, or 8), at year 10, and at the end of the study (typically from year 10 to year 13).

- Even if a participant chose to discontinue the treatment during the clinical trial, study treatment (FOLLTRT) was documented as present. The randomized study treatment could change for a participant during the clinical trial due to reassignment (for smokers wishing to not take beta-carotene) or subsequent to the clinical trial when participants were free to take the study ingredients of their choice.

- Although dates of occurrence are suppressed for reasons of confidentiality, select event times (from the date of enrollment), in years, are made available, including: study visit time, time to cataract surgery (if applicable), time to photocoagulation for neovascular AMD, and approximate time of follow-up interview (for those visits where the follow-up interview occurred).

- EOTVIS: This field is "Y" for the record that represents the last study visit prior to the end of the clinical trial; otherwise, it is missing.

- ARTHRITIS, GOUT, DIABETES, GLAUCOMA, GLAMEDS, GLASRGRE, GLASRGLE, ANYEYERE, ANYEYELE, CAPSULRE, CAPSULLE, REPRETRE, REPRETLE, OCUSRGRE, OCUSRGLE, VOTSRGRE, VOTSRGLE, LNSPRSRE, LNSPRSRE: The questions that yielded data in these fields were only asked at annual visits in Phase II and at in-clinic visits in Phase III.

- GLAMEDS: This field is "Y" if question 10a on the annual visit form is "yes" for either the right eye or left eye.

- LE_ERR_F, RE_ERR_F: Only measured at non-annual visits in Phase II when Chart R visual acuity loss from baseline was 10 or more letters in either eye.

- FOLLVALE, FOLLVARE: When measured at non-annual visits in Phase II, these fields contain Chart R values if visual acuity loss from baseline in both eyes was no more than 9 letters; if Chart R visual acuity loss from baseline was 10 or more letters in either eye or the visit was an annual visit or an in-clinic visit, the fields contain Chart 1 & 2 values from best-corrected vision.

- IOPRE, IOPLE: These data were measured only at non-annual visits in Phase II and at in-clinic visits in Phase III.

- HYDROCH5, DIPYRID5 TRIAMTE5, SPIRONO5, PROPRAN5, BETABLO5, CHLORTH5, FUROSEM5, ACEINHI5, CALCBLO5, POTASSI5, DIGOXIN5, NITROGL5, ISOBDIN5, ALLOPUR5, BENEMID5, BLOODTH5, LOWRCHO5, ANTIINF5, THYRHOR5: These medication use measures indicate number of years taken for any medication that the participant is currently taking at that study visit.

1.1.2.2 NSAIDs Use (all participants) [table name: nsaids]

Historical use of NSAIDs was collected once from most participants, but sometimes twice, during a follow-up visit within the course of only one calendar year during the Natural History portion of AREDS.

- Time (in years) between date of enrollment and date of NSAIDs data collection is included (NSAIDSTIME).

1.1.2.3 Fundus Photographs (all participants) [table name: fundus]

This table contains fundus photograph grading data only (no images) for all participants from the AREDS Photograph Reading Center for all study visits where photographs were taken, including the baseline (00) visit. With the following exceptions, all grading data has been directly mapped from fields on the Maculopathy Grading form.

REELICAT, LEELICAT:	Eye-based AMD Category.
DRUCAT:	Participant-based AMD Category.
AMDSEVRE, AMDSEVLE:	Eye-based AMD Severity Scale Score. See AREDS Report No. 17. *Arch Ophthalmol.* 2005 Nov; 123(11):1484-98.
SCALE:	Participant-based Simple Scale Score. See AREDS Report No. 18. *Arch Ophthalmol.* 2005 Nov; 123(11):1570-4.

Table details are as follows:

- Fields designating the right eye start with RE; fields designating the left eye start with LE.

- Time (in years) between date of enrollment and date of the photograph are included for each eye (REPHTIME, LEPHTIME).

- Photographs were taken during follow-up at annual visits starting with visit 04 (year 2). Thus, most visit numbers are even.

- Photographs that were missed at an even-numbered visit (e.g., 06) were then taken at the next odd-numbered visit (e.g., 07), but nonetheless assigned the prior even numbered visit (i.e., 06).

- If a photograph visit number is odd (01, 03, etc.) or 02, this photograph was taken due to a participant's vision loss compared to baseline and is considered an "event" photograph. See sections 6.4.1 and 6.4.2 of the AREDS Manual of Operations for more information.

- Unless otherwise noted in the data dictionary, a value of '8' indicates that a grade could not be obtained for that measurement.

- REELICAT, LEELICAT (Eye based AMD Category) and DRUCAT (Participant-based AMD Category) are fields that are not visible on the grading form, but were included as summary measurements in the original AREDS data. These categorizations are consistent with the AMD Categorization found in the AREDS Manual of Procedures section 3.1.2.8, with the exception that visual acuity is not considered.

- DRUCAT: Participant category based on worse of two eyes, if different. Categories 3a and 3b were only used at the baseline (00) visit. Refer to the AREDS Manual of Procedures, section 3.1.2.8, for an explanation of these categories.

- The circles (I-2, O-2, etc.) represent a set of graduated measurement circles for estimating the area involved for various abnormalities (Geographic Atrophy, RPE Depigmentation, Increased Pigment, Maximum Drusen Size, and Drusen Area). These areas can be converted into disk diameters (DD) or microns (μ) as follows:

  C-0: 0.042 DD/63 μ
  C-1: 0.083 DD/125 μ
  C-2: 0.167 DD/250 μ
  I-1: 0.120 DD/175 μ
  I-2: 0.241 DD/350 μ
  O-1: 0.219 DD/322 μ
  O-2: 0.439 DD/644 μ

  More explanation on these measurements can be found in the following AREDS publications:

  - AREDS Report No. 6 – The Age-Related Eye Disease Study System for Classifying Age-related Macular Degeneration from Stereoscopic Color Fundus Photographs. *American Journal of Ophthalmology*. Vol. 132 (No. 5), November 2001. pp 668-81.
  - AREDS Report No. 17 - The Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration. *Archives of Ophthalmology*. Vol 123, November 2005. pp 1484-98.

- Drusen area (REDRARWI, LEDRARWI) is missing if a participant had advanced AMD in that eye.

- AMDSEVRE, AMDSEVLE: This is the AMD Severity Scale score in each eye that was calculated based on fields from the Maculopathy Grading form. More details on the calculation of these fields can be found in AREDS Report No. 17.

- SCALE: This is the AMD Simple Scale score that was calculated based on fields from the Maculopathy Grading form. The score is a person-based score derived based on the drusen/pigment status in each eye. This score ranges from 0 to 4 at baseline and 0 to 5 over follow-up. A score of 5 is assigned to participants with advanced AMD in both eyes. This is a modification of the scoring method described in AREDS Report No. 18 - A Simplified Severity Scale for Age-Related Macular Degeneration. *Archives of Ophthalmology*. Vol 123, November 2005. pp 1570-74.

- The visit that represents the last photograph taken during the clinical trial has a “Y” value for LASTCLINF. All visits subsequent to this visit represent photographs taken during the natural history portion of AREDS.

- The first visit after the visit where LASTCLINF is Y may be odd-numbered (e.g., 13, 15, etc.) based on the timing between the first AREDS visit and the start of the natural history portion of AREDS. Therefore, if these odd visit photographs occur

during the natural history portion of AREDS, they are not necessarily "event" photographs.

1.1.2.4 Lens Photographs (all participants) [table name: lens]

This table contains lens photograph grading data only (no images) for all participants from the AREDS Photograph Reading Center for all study visits where photographs were taken, including the baseline (00) visit. With the following exception, all grading data has been directly mapped from fields on the Detailed Lens Grading form:

RENUCADJ, LENUCADJ:	Nuclear sclerosis adjusted from RENUCSCL and LENUCSCL, respectively, to fit onto a scale from 0.9 to 6.1, whereas RENUCSCL and LENUCSCL were graded on a scale from 0.9 to 7.1.

Table details are as follows:

- Time (in years) between date of enrollment and date of the photographs are included for each eye (RENZTIME, LENZTIME for Neitz photographs; RESLTIME, LESLTIME for Slit Lamp photographs).

- Photographs were taken during follow-up at annual visits starting with visit 04 (year 2). Thus, most visit numbers are even.

- Photographs that were missed at an even-numbered visit (e.g., 06) were then taken at the next odd-numbered visit (e.g., 07), but nonetheless assigned the prior even numbered visit (i.e., 06).

- If a photograph visit number is odd (01, 03, etc.) or 02, this photograph was taken due to a participant's vision loss compared to baseline and is considered an "event" photograph. See sections 6.4.1 and 6.4.2 of the AREDS Manual of Operations for more information.

- Unless otherwise noted in the data dictionary, a value of '8' indicates that a grade could not be obtained for that measurement.

- REPCTCOA, LEPCTCOA (Cortical Opacity %: whole grid), REPCTCOL, LEPCTCOL (Cortical Opacity %: central 5 mm circle), REPCTPSC, LEPCTPSC (Posterior Subcapsular Opacity %: central 5 mm circle) are not fields that are visible on the grading form, but were included as summary measurements in the original AREDS data.

- The visit that represents the last photograph taken during the clinical trial has a "Y" value for LASTCLINL. All visits subsequent to this visit represent photographs taken during the natural history portion of AREDS.

- The first visit after the visit where LASTCLINL is Y may be odd-numbered (e.g., 13) because of when the participant had their first visit during the natural history portion of AREDS and is not necessarily an "event" photograph.

1.1.2.5 Visual Functioning Questionnaire [VFQ-25] [table name: vfq]

This table contains scoring data from the Visual Functioning Questionnaire (VFQ).

Table details are as follows (field names or question numbers from VFQ form shown in parentheses below):

- This questionnaire was collected at the following time points: Year 4 or earlier (for participants who were not beyond year 4 when the questionnaire was implemented), year 5, end of the clinical trial, year 10, and end of the study.
- Time (in years) between date of enrollment and date of questionnaire administration is included (VFQTIME).
- GENH (General Health Subscale Score): Combination of responses to Overall Health (1) and GENERAL HEALTH subscale (A1).
- GENVIS (General Vision Subscale Score): Combination of responses to eyesight (2) and GENERAL VISION subscale (A2).
- OCPAIN (Ocular Pain Subscale Score): Combination of responses to pain/discomfort around eyes (4) and limitations due to pain in/around eyes (19).
- NEARACT (Near Activities Subscale Score): Combination of responses to difficulty reading newspapers (5), doing work requiring that a person sees well up close (6), finding something on a crowded shelf (7), and NEAR VISION subscale (A3, A4 & A5).
- DISTACT (Distance Activities Subscale Score): Combination of responses to difficulty reading street signs or store names (8), going down steps in dim light (9), going to see movies, plays, or sports events (14), and DISTANCE VISION subscale (A6, A7, & A8).
- VSOCFUNC (Vision Specific: Social Functioning Subscale Score): Combination of responses to difficulty seeing how people react (11), visiting with people (13) and SOCIAL FUNCTION subscale (A9).
- VMENTALH (Vision Specific: Mental Health Subscale Score): Combination of responses to being worried about eyesight (3), frustration due to eyesight (21), loss of control due to eyesight (22), worry over embarrassment due to eyesight (25), and WELL-BEING/DISTRESS subscale (A12).
- VROLEDIF (Vision Specific: Role Difficulties Subscale Score): Combination of responses to accomplishing less because of vision (17), limited in activities because of vision (18), and ROLE LIMITATIONS subscale (A11a & A11b).
- VDEP (Vision Specific: Dependency Subscale Score): Combination of responses to staying home most of the time (20), relying too much on other people (23), needing a lot of help from others (24), and DEPENDENCY subscale (A13).

- DRIVE (Driving Subscale Score): Combination of responses to about driving a car (15), driving at night (16), and DRIVING subscale (A10). This score is missing if the participant noted that they did not drive.

- AVTOT (VFQ Overall Score): Total scale score computed by averaging the sum of the subscales. Note, if the participant did not drive, the average excludes the DRIVE subscale.

1.1.2.6 Dietary Information [table name: dietary]

This table contains nutritional analysis data derived from the AREDS Food Frequency Questionnaire, administered at enrollment and at an additional time point during follow-up.

Table details are as follows:

- Visit numbers (VIS) are either "Baseline" or "Follow-up". The field name differs from the VISNO field found in many of the other tables to avoid confusion.

- Time (in years) between date of enrollment and date of completion of the follow-up Food Frequency Questionnaire is included (FFQTIME). The baseline food frequency questionnaire is noted as missing, but technically speaking may be considered completed at time 0.

- With the exception of dietary glycemic index fields, all dietary values were measured using a scoring system developed at the University of Minnesota as described in the procedures section of the following manuscript:

  - AREDS Report No. 22 – The Relationship of Dietary Carotenoid and Vitamin A, E, and C Intake with Age-Related Macular Degeneration in a Case-Control Study. *Epidemiology*. Vol 125 (No. 9), September 2007. pp 1225-32.

- Dietary glycemic index fields were obtained from research done by Tufts and were only calculated for the baseline record. Please review the section entitled "Assessment of carbohydrate variables" in the following manuscript:

  - Chung-Jung Chiu, Roy C Milton, Gary Gensler, and Allen Taylor. Dietary carbohydrate intake and glycemic index in relation to cortical and nuclear lens opacities in the Age-Related Eye Disease Study. *Am. J. Clinical Nutrition*, May 2006; 83: 1177 - 1184.

1.1.2.7 Adverse Experiences (all participants) [table name: adverse]

This table contains all reported adverse experiences that may have been related to the study treatment.

Table details are as follows:

- There is no visit number field for this table.

- Adverse experience reporting occurred for the duration of the AREDS clinical trial and for the first 6 months of the natural history portion of AREDS.

- Time (in years) between date of enrollment and date the adverse experience was first observed is included (OBSTIME).

- Time (in **days**) between date the adverse experience was first observed and date of resolution is included (ADVTIME). This field may be missing if the adverse experience was considered chronic or was never resolved.

- The Adverse Experience Type (ADVTYPE) serves as the primary description of the adverse experience. When the type was "Other" (values of 50-99) or warranted additional detail (a value of 28 or 33), the ICD-9 code (ICD9COD1) for that adverse experience is shown. In the three instances where no ICD-9 code was available, the clinical description is shown. (ADVDESC).

1.1.2.8 Hospitalization [table name: hospitalization]

This table contains all reported inpatient hospitalizations for the duration of the study.

Table details are as follows:

- There is no visit number field for this table.

- Time (in years) between date of enrollment and date of admission in to the hospital is included (ADMTIME).

- Time (in **days**) between date of admission and the date of discharge is included (HOSPTIME).

- The primary ICD-9 code (ICD9COD1) based on the discharge summary serves as the primary hospitalization cause. When the reason (REASON) was "Other" and no ICD-9 code was available or the discharge summary was not provided, the clinical description (REASONSP) is shown.

1.2 Phenotype Data Tables for Participants with a Genetic Specimen

These two tables provide phenotype data for participants for whom at least one genetic specimen was collected. The reason these tables are needed is that some participants who submitted a genetic specimen consented that it be used for Eye Disease Research Only, whereas some data in the Project Data Tables could be used for other types of disease research.

- The ID2 field represents the unique participant identifier, whereas the ID field serves this function in the Project Data Tables.

- For participants who consented to General Research Use for their genetic specimen, the value of the ID2 field is equal to the value of the ID field. This means that the data in the Phenotype Data Tables can be linked to the data in all Project Data Tables using the ID field.

- For participants who consented to Eye Disease Research Only, the value of the ID2 field is not the same as the ID field; it is a 4-digit code that begins with the letter 'G' followed by 3 numbers. This means that the data in the Phenotype Data Tables can NOT be linked to the data in the Project Data Tables.

- This design means that the phenotypic information available for genetic specimens from participants who consented to Eye Disease Research Only is limited to a much smaller set of variables considered pivotal for eye disease research but not considered pivotal for other types of disease research. By contrast, genetic specimens from participants who consented to General Research Use can be linked with all available data from the Project Data Tables.

The description of genetic specimens as 'Matched' or 'Untested' is included because during a pilot examination of participants from one AREDS Clinical Center with 2 different blood samples drawn, it was found in some cases that the two samples presumed to be from the same participant did not match. To ensure that this labeling error was not more widespread, a second blood sample was obtained from most participants and the DNA fingerprint from both blood samples was compared.

- A 'Matched' specimen means this participant had 2 or more different blood specimens whose DNA fingerprints matched. Of the participants with multiple specimens whose DNA fingerprints were compared, 2.9% had specimens that did not match. Specimens that were DNA fingerprinted but did not match were destroyed.

- An 'Untested' specimen means either of the following:

    - No second sample was obtained for this participant so the matching status could not be tested (the majority of cases).
    - A second sample was obtained for this participant but DNA was not obtained for fingerprinting. However, other genetic material (such as a cell line) is available that could be used to obtain DNA in the future which could then be tested. This means that in some rare cases a participant has both 'Matched' and 'Untested' specimens.
    - Based on the DNA fingerprinting results described above, it is estimated that 2.9% of untested specimens may have a labeling error such that the AREDS ID of the genetic specimen has been linked to the phenotypic data from a different AREDS participant.

DNA from AREDS participants can be obtained for use in disease research from the NEI-AREDS Genetic Repository which is located at the Coriell Institute for Medical Research (http://www.coriell.org). The ID2 variable is utilized for the purpose of finding specimens of interest. Note that not all AREDS genetic specimens have DNA available.

1.2.1 Eye Disease Phenotypes [table name: genspecphenotype]

Includes one record for every AREDS participant for whom at least one genetics specimen was collected, using ID2 as the key identifier. This data table contains information about participant phenotype (AMD and lens), AMD and lens status at select time points, selected demographic variables at the time of the participant's enrollment, as well as annual health status measurements for smoking, body-mass index, blood pressure, angina, cancer, and diabetes.

Table details are as follows:

- Data for health status measurements are missing for participants whose ID2 field starts with a G, as these participants only consented to research relating to eye disease, not general health.

- The following fields are taken directly from the Enrollment/Randomization Project Data Table: AMDCAT, CNTRMTAK, ENROLLAGE, MARITAL, RACE, SCHOOL, SEX, TRTCAT, DIABAGE, DIABDIET, DIABINS, DIABPILL.

- The following fields are taken from the Eye Disease Phenotypes Project Data Table: AMDSTAT, CATARACT, COR, LCOR, RCOR, LCORBASE, RCORBASE, LCORSCORE, RCORSCORE, NUC, LNUC, RNUC, LNUCBASE, RNUCBASE, LNUCSCORE, RNUCSCORE, PSC, LPSC, RPSC, LPSCBASE, RPSCBASE, LPSCSCORE, RPSCSCORE.

- CNTRMTAK: Collected at baseline, this indicates the participant's intention to supplement with Centrum during the study.

- Annual health status measurements span from baseline (field names ending in '00') through year 13. Cancer and diabetes information was collected at least annually. Angina, smoking status, body-mass index, and blood pressure were collected at baseline, year 5, end of the clinical trial (typically year 6, 7, or 8), year 10, and the end of the study (typically from year 10 to year 13) and thus there are some years (3, 4, 9, e.g.) where there is limited information on these measurements.

- The SMK00 and SMK03 through SMK13 fields were calculated as shown below using two questions, 1) whether they had ever smoked for at least 6 months and 2) if they are a current smoker.

  - 'Never smoked': Ever smoked 6 months = N
  - 'Former smoker': Ever smoked 6 months = Y, Current smoker = N
  - 'Current smoker': Ever smoked 6 months = Y, Current smoker = Y

- SYST00, DIAS00, SYST03 through SYST13, and DIAS03 through DIAS13: Systolic and diastolic blood pressure data included are from the second reading.

- RNUCBASE, LNUCBASE: The data shown are the same as RENUCADJ and LENUCADJ from the Fundus & Lens Photographs Phenotype Data Table (see below).

- RCORBASE, LCORBASE: The data shown are the same as the baseline value shown for REPCTCOA and LEPCTCOA from the Fundus & Lens Photographs Phenotype Data Table (see below).

- RPSCBASE, LPSCBASE: The data shown are the same as the baseline value shown for REPCTPSC and LEPCTPSC from Fundus & Lens Photographs Phenotype Data Table (see below).

1.2.2 Fundus & Lens Photographs [table name: genspecfunduslens]

This table contains combined data from the Fundus Photographs and Lens Photographs Project Data Tables (see detail, above, for each of these two tables), but only includes AREDS participants for whom at least one genetic specimen was collected and ID2 is the key identifier. As with the Fundus Photographs and Lens Photographs tables, there are multiple records per participant.

2. Analysis of AREDS Data

2.1 Characterizing the Cohort

2.1.1 Randomized Treatment (from Enrollment/Randomization table)

- The AREDS treatment assignment values (TRTCAT) are as follows:
  - 1: Placebo
  - 2: Antioxidants only
  - 3: Zinc only
  - 4: Antioxidants+zinc

- Participants in AMD Category 1 at randomization were randomly assigned with equal likelihood to Placebo or Antioxidants, as they were unlikely to progress to advanced AMD and thus were not put at potential additional risk by supplementing with high-dose zinc. Participants in all other AMD Categories were randomly assigned with equal likelihood to any of the four treatments.

- All participants except for those in AMD Category 1 were included in the AMD Clinical Trial. In this trial, the main effects models (the two Antioxidant-containing groups vs. the two non-Antioxidant-containing groups, or the two Zinc-containing groups vs. the two non-Zinc-containing groups) could be analyzed as well as the effect of each individual arm vs. placebo.

- All participants without a history of bilateral cataract surgery at baseline were included in the Cataract Clinical Trial, but only the main effects model of the two Antioxidant-containing groups vs. the two non-Antioxidant-containing groups provided an unbiased, balanced analysis. (No category 1 patients were assigned to zinc and including them in a Zinc vs. no Zinc analysis is inappropriate because of known confounding factors and such an analysis is not a randomized comparison.)

- Detailed information about the inclusion/exclusion criteria can be found in the AREDS Manual of Operations, section 3.1.3.

- As noted previously, Centrum® usage (or intent) during the study was typically analyzed by using the CNTRMTAK field, as Centrum adherence was not specifically documented during the study.

2.1.2 Clinical Visual/Ocular Characteristics (from Enrollment/Randomization table)

- History of cataract surgery in each eye (CATSRGQR, CATSRGQL) and history of laser photocoagulation for NV AMD (PRVLASQR, PRVLASQL) can be used to determine whether it is appropriate to include an eye in longitudinal analyses.

- The baseline visual acuity data for each eye are found in the ACUIT1RR (Chart 1, right eye) and ACUIT2RL (Chart 2, left eye) fields.

- Baseline participant AMD Category is found in the AMDCAT field. This category was determined by a combination of the qualification visual acuity (QUALVARE, QUALVALE) and the AREDS Photograph Reading Center grading of each eye (see the detail about REELICAT and LEELICAT, below). Although categories 3a and 3b (as well as 4a and 4b) represent participants meeting slightly different eligibility criteria, these individuals can sometimes be collapsed into a single category 3 or 4, respectively. On other occasions, it may be necessary to exclude AMD Category 3b and AMD Category 4b participants from analyses. Please refer to the AREDS Manual of Procedures, section 3.1.2.8, for an explanation of these categories.

- When doing analysis by eye as opposed to by participant, note that participants in AMD Categories 3b, 4a, and 4b only have one eligible study eye (see detail below for how to identify the study eye).

2.1.3 Eligible Study Eyes for AMD Clinical Trial (from Fundus Photographs table)

- Study eye eligibility is determined using baseline REELICAT (right eye AMD Category) and baseline LEELICAT (left eye AMD Category).

- Study eye eligibility is as follows:

  - Eyes graded as 1, 2 or 3 at baseline are eligible study eyes.
  - Eyes graded as a 4 at baseline have advanced AMD and are not considered study eyes.
  - Eyes graded as a 5 had a lesion disqualifying that eye from the study.
  - A few eyes were graded as 8 meaning their photographs were not gradable and are also not included as “study” eyes.

2.1.4 Photographic Ocular Characteristics (from Fundus or Lens Photographs table)

Specific ocular baseline characteristics can be found in the *fundus* or *lens* tables in the record where VISNO = 00. The first four items refer to fields found in the *fundus* table and the last item to fields in the *lens* table.

- Presence of Geographic Atrophy (GA) at baseline is defined as having a value of 2 through 7 for REGEOAWI (right eye) or LEGEOAWI (left eye).

- Presence of central GA is indicated by either having a value of 2 for REGEOACT (right eye) or LEGEOACT (left eye); Central GA may also be indicated by having a value of 1 for REGEOACT and a value of 2 through 4 for REGEOACS or a value of 1 for LEGEOACT and a value of 2 through 4 for LEGEOACS.

- Drusen size (REDRSZWI or LEDRSZWI) was instrumental in determining AREDS participant baseline AMD Category (AMDCAT). Although drusen area (REDRARWI or LEDRARWI) was not utilized in determining AMD Category, it was useful in determining AMD Severity Scale Score (AMDSEVRE or AMDSEVLE).

- Cataract status was determined by nuclear sclerosis (RENUCADJ or LENUCADJ), cortical opacity (REPCTCOL or LEPCTCOL) or PSC opacity (REPCTPSC or LEPCTPSC). RENUCSCL, LENUCSCL, REPCTCOA and LEPCTCOA were used less frequently in analyses. RENUCADJ and LENUCADJ, the rescaled values of RENUCSCL and LENUCSCL, were intended to condense milder nuclear sclerosis into fewer steps on the scale. REPCTCOL and LEPCTCOL measure cortical opacity within the 5mm circle, whereas REPCTCOA and LEPCTCOA measure cortical opacity within the grid and may reflect non-severe opacities.

2.1.5 Dietary Characteristics

Dietary measurements at baseline may also be included in the analysis by selecting the 'baseline' record (VIS) from the *dietary* table.

2.2 Determining Outcomes

2.2.1 Ocular Outcomes

Although not essential, it is recommended that the Final AMD Phenotype and Final Cataract Phenotype measures described in the Eye Disease Phenotypes data tables are used, as these data represent a thorough evaluation and categorization of AMD and cataract by AREDS researchers who had the most intimate understanding of the data. Furthermore, consistent use of these outcome measures across all investigators utilizing this database will allow for greater opportunity to compare results.

For investigators who wish to evaluate the longitudinal data for ocular outcomes, please note the following:

- When evaluating a longitudinal AMD outcome, ensure that you are only analyzing eligible study eyes (see "Eligible Study Eyes", above).

- When evaluating a longitudinal cataract outcome, ensure that you are only analyzing eyes that did not have a history of cataract surgery.

- Progression to advanced AMD was defined in AREDS as follows:

  - A value of '4' for the REELICAT or LEELICAT fields from the *fundus* table starting at year 2 (visit 04), OR
  - Laser photocoagulation for AMD whereby the date of laser photocoagulation (AMDTRTRE, AMDTRTLE) from the *followup* table is not missing (this can occur any time after randomization).

- It is possible that advanced AMD was noted at one study visit and was not seen at one or more successive study visits. This occurrence, among others, led the AREDS researchers to develop the more comprehensive Final AMD Phenotype, available in the Eye Disease Phenotypes table.

- Photograph data from the fundus table at visits 01 through 03 are sparse, as photographs were only taken at these visits in conjunction with a vision loss or when the development of advanced AMD was observed by the clinician.

- Longitudinal cataract events, including nuclear sclerosis progression and increase in cortical or PSC opacity, often could not be properly analyzed due to lens removal. The Final Cataract Phenotype in the Eye Disease Phenotypes table provides the best assessment of each type of cataract opacity, plus an overall assessment of which opacity, if any, was most influential in cataract progression.

2.2.2 Other Non-ocular Outcomes

The AREDS Project Data Tables, in particular the *followup* and *hospitalization* tables, provide an opportunity to evaluate other non-ocular outcomes. Please read through the detail of the Project Data Tables carefully when evaluating non-ocular outcomes, as some data points may not be measured regularly which may lead to bias in the analysis.

2.2.3 Clinical Trial vs. Natural History Phases of AREDS

The AREDS Clinical Trial phase (considered Phase II) began in November 1992 and was completed in October 2001. About 80% of AREDS subjects alive at the time the Clinical Trial ended chose to continue being seen as part of the Natural History phase (considered Phase III), which started in October 2001 and was completed by December 2005. To provide precise indications as to which study data were collected during the Clinical Trial phase and which study data were collected during the Natural History phase, please utilize the following data fields (please see table detail for more information):

followup:	EOTVIS
fundus:	LASTCLINF
lens:	LASTCLINL

Data from the *mortality*, *hospitalization*, *adverse*, and *vfq* tables were collected during both phases. Please compare the “time” field from each of these tables (DEATHTIME, ADMTIME, OBSTIME, and VFQTIME, respectively) to the VISTIME field in the *followup* table to determine during which phase the data record occurred. All data from the *dietary* table were collected during the Clinical Trial phase. All data from the *nsaids* table were collected during the Natural History phase.